Developing an Intelligent Human-Computer Interaction System Using Deep Learning and Machine Learning Algorithms

Dr. P. Balaji Srikaanth[1], Dr. Santhosh Kumar Chenniappanadar[2]
[1]Associate Professor, Dept. of Networking and Communications, School of Computing,
College of Engineering and Technology, SRM Institute Science and Technology,
Kattankulathur, Chennai-India.
[2]Department of Information Technology, Sona College of Technology,
Salem-636005, Tamil Nadu, India.

Article Info

Article history:

Received Mar 12, 2025
Revised Apr 10, 2025
Accepted May 11, 2025

Keywords:

Deep learning
Human–Machine Interfaces (HMIs)
Gesture recognition
Speech recognition
Human-computer interface

ABSTRACT

As crucial input techniques in Human-Computer Interaction (HCI), speech and gesture recognition have become increasingly popular in virtual reality in recent years. Specifically, the swift advancement of deep learning, artificial intelligence, and other computer technologies has led to revolutionary advancements in voice and gesture detection. Significant technical advancements in the field of Human-Computer Interaction (HCI) have made it possible for educators to deliver high-quality educational services by utilizing intelligent input and output channels. This work proposes a straightforward and efficient way for extracting salient characteristics based on contextual information by conducting in-depth research and analysis on the design of human-computer interaction systems using machine learning algorithms. The findings demonstrate how deep learning and intelligent HCI are extensively used in voice, gesture, emotion, and intelligent robot direction. In related study disciplines, a wide range of recognition techniques were put forth and experimentally validated. When it comes to voice-activated Human–human-machine interfaces (HMIs), context is crucial to enhancing user interfaces. Action recognition accuracy and precision can be significantly increased by combining long short-term memory networks with convolutional neural networks. As a result, more industries will be involved in the use of HCI in the future, and more opportunities are anticipated.

Corresponding Author:

Dr. P. Balaji Srikaanth,
Associate Professor, Dept. of Networking and Communications, School of Computing,
College of Engineering and Technology, SRM Institute Science and Technology,
Kattankulathur, Chennai-India.
Email: balajis7@srmist.edu.in

1. INTRODUCTION

The field of study known as HCI examines how people create and utilize PCs and technology [1]. Modern information and communication technologies have advanced so quickly that computers now serve a very different purpose. Computers have become essential for daily living, work, social interactions, and education. Computer programs must have user interfaces with an effective HCI, regardless of the device type (e.g., desktop, laptop, smartphone, etc.). This will allow users to efficiently complete a variety of tasks, including typing documents, operating a vehicle, controlling a robot arm, conducting internet searches, listening to music, and more the widespread use of technology in today's culture and its pervasiveness in our day-to-day activities have made human-computer interaction (HCI) one of the more fascinating research areas in recent years. Many researchers have created clever and creative ways to make HCI appealing to users as a result of the demand for practical HCI technology.

AI encompasses a several automated methods for making decisions, such as neural networks and conditional logic [2]. ML, a subset of AI, refers to decisions or predictions produced using data-driven methodologies. DL, as shown in Figure 1, is the subset of ML methods that make use of DNN. AI research articles now make up 3% of all published journal articles and 9% of all convention publications during the past 20 years. Most of this study of AI usually focuses on creating algorithms and improving technology, with a particular focus on accurate and performant models. In addition to scholarly study, ML, AI, and DL are widely used in data-rich sectors. With differing degrees of success, these same businesses have developed goods and services on top of an AI back-end. Analyzing and improving the process of translating from the computational model to end-user requirements is becoming more and more necessary as the industry continues to integrate performant models.

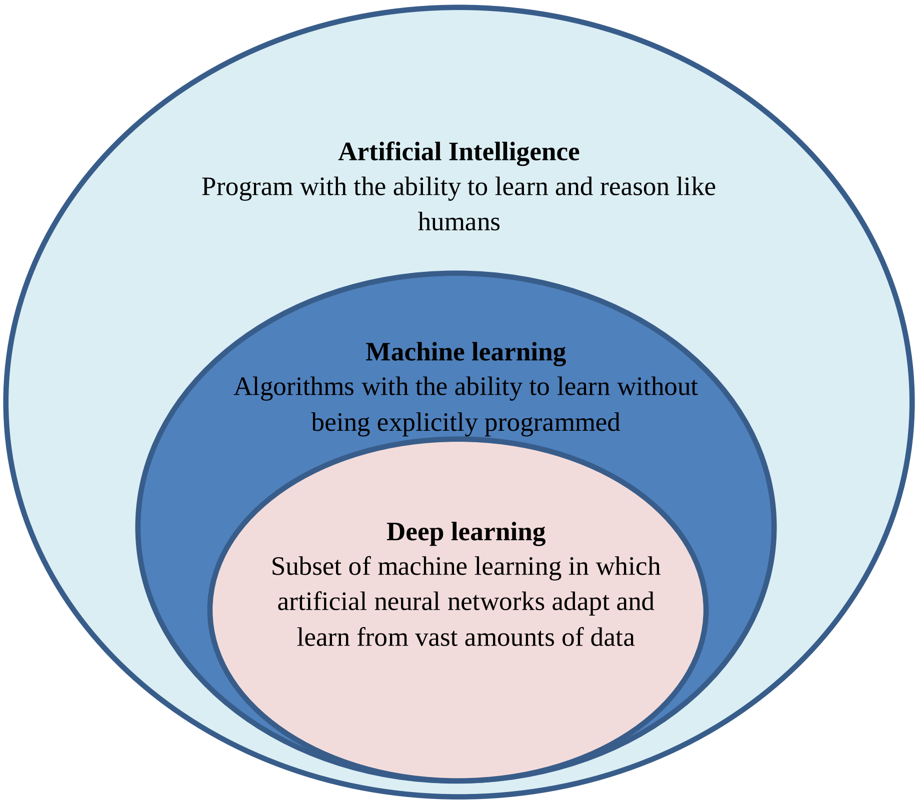

Figure 1. Overview of Artificial Intelligence

The technology that allows people to use computers is known as the HCI. How we can simplify computer interaction so that it is more advantageous for people with disabilities [3]. One fresh innovation that will be utilized to provide an interface for computer system operation is HCI. These days, there are many other ways to get attention, such as through internet browsers, desktop applications, and portable electronics. These user interfaces are graphical. Voice synthesis and recognition are more applications for voice interface systems. Gray matter, white matter, and care spinal fluid are used to categorize magnetic resonance images of the human brain. The structural and functional characteristics of the human brain are displayed in these areas. That will provide you with specifics about how to communicate with the machine. The purpose of this region-based study was to identify human-computer interaction.

The role of intelligent HCI systems has important implications concerning accessibility and social inclusion. For instance, for people with mobility disabilities, older users, or those with low capability or willingness to relate to screens (aka limited digital literacy), traditional interface systems are barriers to usability (i.e., keyboard and screen). Instead, users can employ interfaces powered by Deep Learning methods that allow them to interactively use voice commands and gestures (or eyes) to create more intuitive, hands-free interactions to promote better independence and enhanced quality of life. Moreover, with adaptive HCI systems, and computer-based agents, it is possible to tailor interactions based on user behaviours, preferences, and contextual needs - achieving needs or service personalization benefits for diverse populations. Given that societies worldwide are increasingly becoming digital, creating inclusive and userfriendly HCI technologies is critical to ensuring equal access and minimizing digital divides across various populations in different settings.

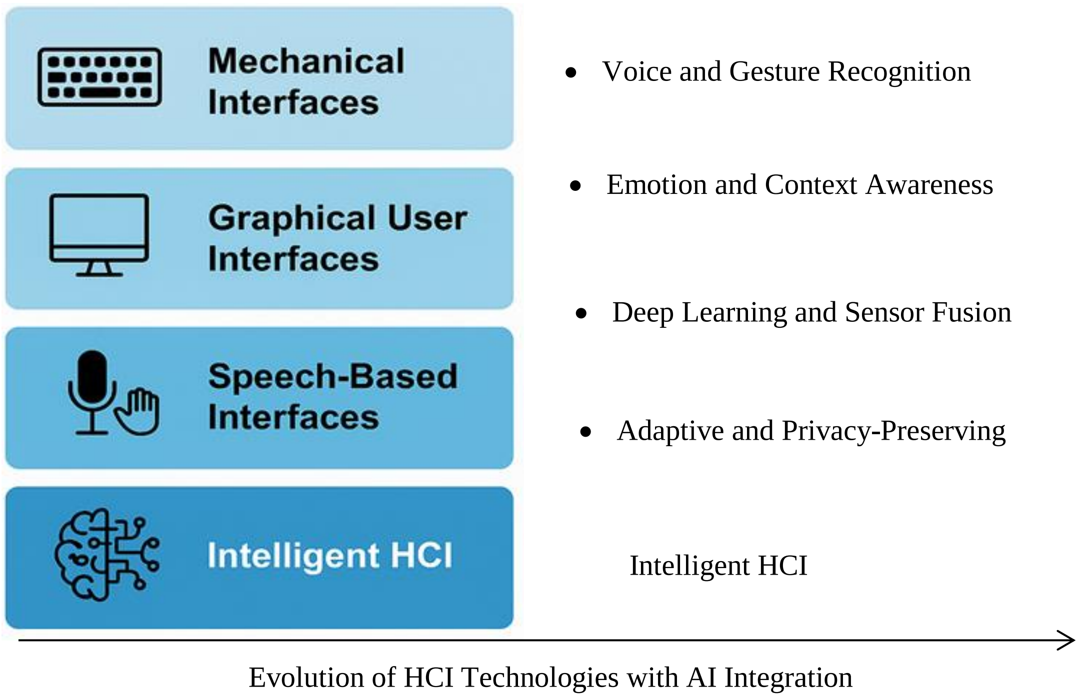

Figure 2. Evolution of HCI Technologies with AI Integration

Figure 2 depicts a chronological evolution of HCI technologies, which began with the originally purely mechanical interfaces and transitioned itself into intelligent HCI systems that now offer various options for sophisticated capabilities, including voice and gesture recognition, emotionally-aware responses, deep learning-powered adaptations, and privacy-preserving personalization. This transition has been driven by greater demand for more natural, aware, and accessible user experiences across industries.

Recent advancements in ubiquitous computing and ambient intelligence have changed the focus of HCI systems by going beyond the basics of response to the user’s explicit intentions, into anticipating needs with implicit understanding or contextual inference and real-time behavioral learning. The explosion of multi-sensor fusion designs that are utilizing EEG, EMG, and eye-tracking data, is taking HCI from being a reactive tool to a proactive assistant. This newly emerging paradigm of HCI highlights the need to design ethically intelligent, inclusive and adaptable systems to support the needs for diverse user insightful capabilities, such that HCI response to learning is seamless and well thought out from the user's perspective [4].

An additional element of advancement in intelligent HCI is the advent of affective computing, which affords systems the ability to determine user emotional state and react accordingly. This technology is extremely useful for applications in healthcare, education, and mental well-being, where understanding user confusion, frustration, or engagement can enhance the user experience. The ability of the emotion-aware interface, or intelligence augmented system, powered by deep learning, to analyze and respond to users' facial expressions, tone of voice, and physiological signals, allows systems to further adapt their response to creating a more sensitive and engaged user experience.

Simultaneously, the advent of edge computing and lightweight AI models is bringing HCI systems to the endpoint, with the goal of providing faster and more secure interactions, without having to rely on continuous cloud connectivity. This is particularly important for latency-sensitive applications, such as autonomous vehicles, wearables, and assistive technology for differently abled users. At the edge, the combination of AI-driven inference and real-time user feedback guarantees HCI systems remain not only intelligent but also fast, private, and scalable [5].

Combined, these developments signify a change from static, generic user interfaces to dynamic, adaptive ecosystems that focus on personalized user requirements. Future HCI systems will increasingly utilize hybrid AI architectures that incorporate learning, emotion, environment awareness, and ethical stewardship in a human-centric and comprehensive model of interaction.

2. RELATED WORKS

Stacked model information mining is different from standard spatial data mining since it can incorporate results from logging, seismic interpretation, or fine geological reservoir studies in addition to the data itself [6]. Direct use of reservoir data in data mining is inappropriate due to its wide variety of types, dimensions, and potential mistakes. Therefore, the retrieved data should undergo preprocessing, such as coding and merging, processing the physical data's normal distribution, and quality-checking other study findings, before employing a data mining technique. The most crucial step in the data mining procedure is also the one where the technical challenge lies.

One part of a conversation system that handles information exchange between players in human-machine interaction is a dialogue manager. There are two main types of dialogue managers: frame-based conversations and chatbots [7]. In frame-based conversations, experts establish discussion frames with slots and principles that each slot can take, but in chatbots, an agent is Open discussion, as it is generally called, is often taught to function without any knowledge of the dialogue framework. In this work, we primarily focus on chatbots and how they are trained using approaches to management based on based on rules, learning by reinforcement, hierarchy reinforced studying, and sequence-to-sequence training..

The process of making choices of DL techniques must be comprehensible and interpretable in order to evaluate the behaviors of the models, particularly when unexpected or inaccurate results are produced. One of the main factors impeding comprehension is the underlying mathematical models' dimensionality and complexity [8]. To provide better user-specific experiences, sensible personal data is usually continuously gathered, analyzed, and stored via the Cloud. Therefore, privacy must be protected. Although models may experience catastrophic forgetting, such continuous learning for DL techniques is a problem in and of itself, but it is an essential component that must be taken into account to allow for scalability.

AI is becoming a major force behind innovation across a range of industries, like finance, medicine, and education. Applications of AI in HCI have drawn more attention as these technologies advance [9]. By developing systems that comprehend and react to human requirements more naturally, this fusion of AI and HCI seeks to improve user experience. By anticipating user demands, learning from user behavior, and modifying interfaces accordingly, intelligent systems can increase user happiness and usability. A new breed of intelligent systems that can interact with users in previously unheard-of ways has been made possible by the quick development of AI tools such as machine vision and processing of natural languages.

Physical controls, such as shifts, keys, and triggers, are gradually being replaced with virtual controls like icons, pictures, and symbols in modern human-computer interaction interfaces, which have taken over as the primary means of user-system interaction [10]. The structure of many digital pieces is complex, and they are entwined and scattered with one another. According to how the body processes visual information, graphics make up one of the GUI's most instinctive feeling and visual sensations. Interface design heavily relies on visual emotion assessments. It is difficult to accurately measure being a form of sensory interpretation, the individual's need and desire for engagement with emotion meanings.

The multifaceted analysis of power use, which takes the shape of a line chart and histogram, makes it evident how the user's ladder electricity is used as well as the trend in electricity usage over the previous two years [11]. Internet systems, self-service payment, recharging records, manual services, extensive data categories, anti-theft measures, and high-accuracy intelligent metering measures are also included. The best part of this app is that it uses fingerprint recognition to visually analyze power data, allowing residents to see how much electricity is being used. In particular, the software will sound an alarm to remind the user to make their payment on time when their remaining power is low..

The interface between this environment and people is the primary control component. The management part can be thought of as a setting for system collaboration and communication [12]. It is in line with the real state of the environment and is acknowledged by the primary control system. Consequently, the intricacy of a real-time communication system lies not only in guaranteeing the logic's correctness but also in guaranteeing the system's timely accuracy. Second, numerous systems or other simultaneous constructs for target development are required for a real-time messaging system to be completely real-time in order to handle a number of tasks.

Using human voices as input, understanding the emotions they carry, and interacting with emotionally rich people are all ways that speech recognition of emotions in a person-computer dialogue system may transform and intelligently interact with humans [13]. Voice emotion detection systems, on the other hand, can be utilized in contact centers, medical care systems, automotive systems, and other applications to help people deal with real-world problems more rapidly and efficiently. Thus, in human-robot interaction, voice emotion recognition has important theoretical and practical research implications. Both creatures and machines utilize force control for operations like curtain wall installation.

The objective of this chapter and the smart-restorable backspace key project were most similar: to enhance text correction without requiring a lot of backspacing or cursor placement [14]. With the use of this technology, users may swipe the return key to remove text that was in error and then swipe the key again to restore the content after fixing the issue. The method examines the edit distance between the text and the dictionary word to identify error places. The primary drawback of the work is the error detection system, which can only identify misspellings. It is unable to identify the overuse of words or grammatical faults. In contrast, the two experiments in this chapter used deep learning approaches to detect a variety of faults.

The efficiency and accessibility of travel, surroundings, security, administration, service, culture, medical, and other industries in contemporary cities can be enhanced through the creation of the front-end HCI technologies of smart cities [15]. The growth of intelligent urban systems depends heavily on raising the bar for intelligence and improving management in the area of urban operation management. The EMG signal can be utilized to show which muscle states are active, and by analyzing it, one can learn about brain activity. Surface EMG Gesture recognition intelligent HCI has the advantage of being non-invasive, which makes it effective in motion detection, clinical diagnostics, artificial control, and neurological rehabilitation.

3. METHODS AND MATERIALS

3.1 The current state of deep learning adoption in intelligent HCI

The main focus of HCI is on the information flow between computers and human behavior. VR, multimedia, design, and cognitive psychology are all included in the broad field of HCI. The interchange of data in HCI depends on gadgets that interact, such as computer-human and human-computer interaction devices. Examples of HCI devices include a keyboard, mouse, controller, data suit, position tracker, data glove, and pressure pen. Devices that support computer-human interaction include printers, plotters, monitors, screens mounted on helmets, and speakers. The evolution of HCI includes technologies such as voice communication, picture acknowledgment, virtual reality, augmented reality, and tactile interaction, which has become more and more popular in recent years.

Conversation via voice is the most natural communication mode and has the greatest input efficiency of the four technology categories, making it easy to expand product adoption situations. Image recognition is utilized in autonomous driving and security to recognize road conditions and human faces. Immersive experiences for movements, display, and interaction are provided by both AR and VR. Motion sensing enables people to interact with their surroundings or devices by using their body's movements directly, eliminating the need for complex control systems and enabling immersive interaction. As science and technology advance, HCI is evolving as well (Figure 3).

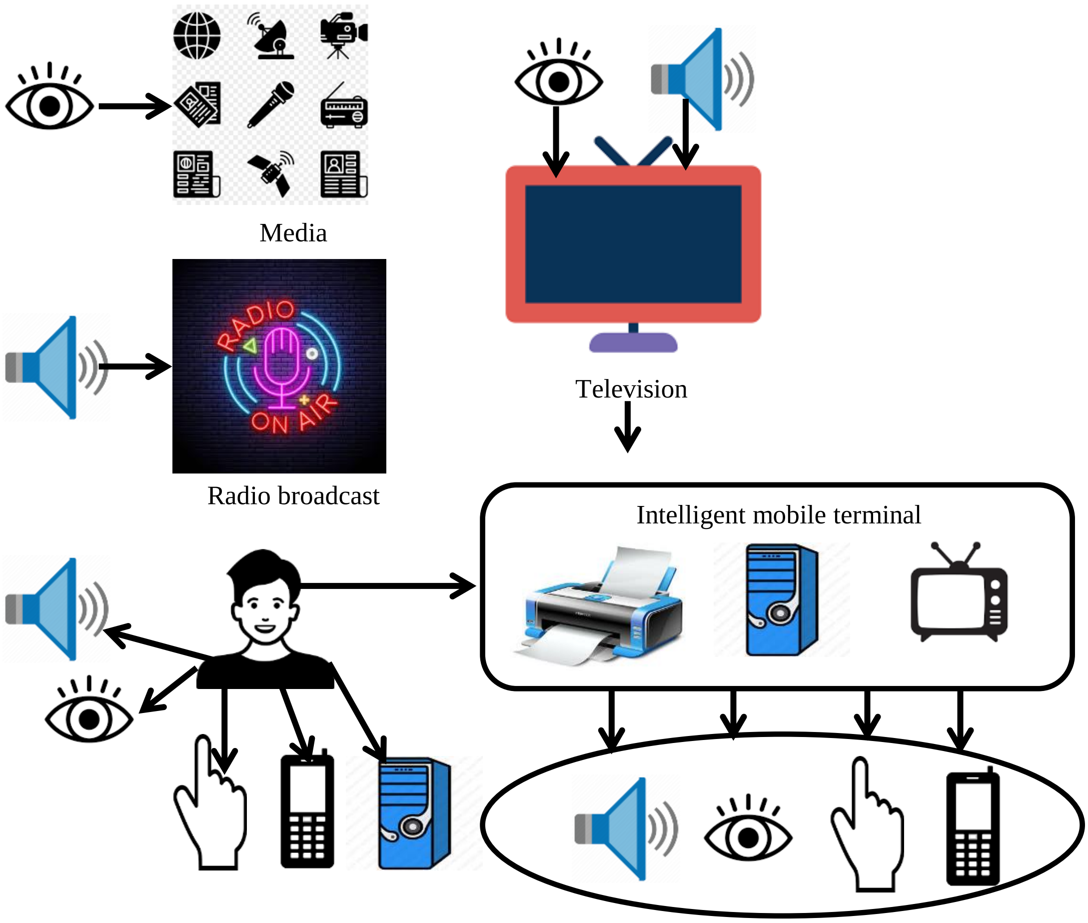

Figure 3. The history of HCI

A recent development in ML research is deep learning. Its growth in speech recognition, information about language processing, and image recognition and retrieval has been comparatively successful in recent years. Additional techniques that employ deep learning to enable human-machine communication include context-aware systems, behavioural creation in user models or incorporated dialog agents, and natural speech processing. The basis for DL adopting is the development of models that mimic the neural connections found in the human brain, which processes sounds, images, and text. After a number of transformation stages, data attributes are displayed hierarchically, and data interpretation comes next.

3.2. Deep Learning's Use in HCI Intelligent Systems

HCI is the process by which people and computers exchange information. This includes the computer giving people data via output or display mechanisms and the people giving the computer pertinent information through input devices. The most important material conveyed by communication behavior in discourse is multimodal simulation, which is the tangible 3D virtual reality of the current environment and coexisting agent. VoxWorld is a modeling platform for creating HCIs that is described in this paper. In a task-oriented collaborative setting, the platform facilitates an integrated dialog system that uses language, gesture, movement, appearance, and gaze tracking to communicate.

The cost of acquiring depth photos is coming down as sensor technology continues to advance. Over time, gesture identification in RGB and depth images has emerged as a field of study in pattern recognition. The majority of deep gesture imaging methods currently in use, however, are really basic, disregard the influence and interaction between the two modes, and do not fully utilize the connected elements between various modes. To address these issues, By considering the independent and connected features of multimodal data, the effect of depth processing on images was maximized and an adaptive weight approach for the fusing of multiple characteristics was established.

Simulation findings showed that the proposed method performed better and had a higher gesture identification rate than the traditional deep gesture image processing method. Furthermore, the proposed method demonstrated its resilience and sustainability by outperforming other advanced gesture recognition systems in terms of identification accuracy. These two research suggest that deep learning-based multimodal image collection can improve HCI systems' gesture detection accuracy.

3.3 Real World Use Case: Intelligent Assistive Wheelchair System

A representative case study of intelligent HCI usage is the development of an assistive smart wheelchair that incorporates speech recognition, gaze tracking and gesture tracking. Utilizing a hybrid CNN-LSTM deep learning algorithm, the system was able to take commands in natural language, such as “move forward” and head gestures to achieve navigation tasks. Eye-tracking, and depth cameras, allowed real time feedback and context aware obstacle avoidance/navigation. Previous work - such as [16] achieved 98.5% accuracy with gaze direction and 1 m/s speeds, and another design that used a CNN-LSTM intent estimation, was able to address false activation issues like the “Midas touch,” and overall increase usability and user trust [17].

3.4 Development Status of Intelligent Voice Interaction System

Voice interface technology powered by AI substantially simplifies several aspects of people's daily life and makes formerly laborious tasks simple and easy to do. The development of the first Individual voice assistance and the current generation of emblematic HCI goods, including intelligent speakers, are examples of AI products that not only enhance people's quality of life but also serve as symbolic representations of scientific and technical advancement. Initially, individuals just used their cell phones to send and receive messages, but these days, they may also be used for voice communication. People only wanted to pay attention to music that pleased them when using traditional speakers.

In addition to providing users with music, 21st-century AI speakers are capable of engaging in a wide range of dialogues with users. In an HCI system, speaking is a crucial and practical technique. It begins by accessing data entered by the user over the whole contact system, which includes data pertaining to voice, face, and multisensory emotions. The dialog system can comprehend the input data, and this dialog section can be used to construct the output later. Lastly, voice synthesis can also display text, with the speech and dialog system components—the entirety of the speech interaction process—being the most crucial components.

People's investigation and advancement demands for AI cannot be satisfied by traditional HCI methodologies, and people's interactions with current AI products are not enough deep. Intelligent speakers, for example, can only meet human and machine needs in a single interaction, but their aesthetic appeal can considerably enhance people's standard of life. Researchers used deep learning to raise the bar for HCI even higher. For instance, to achieve a range of deep HCI mechanisms, deep learning is applied to human-computer communication systems in conjunction with conventional techniques.

3.5 Human Gesture Recognition Based on Deep Learning

One kind of nonverbal expression is a motion, which can be applied in a variety of contexts, including medical applications, robot control, HCI, automation of homes, and communication between the deaf and mute. A variety of methods are used in gesture-based research, like technology for sensors with instruments and visual computing. To put it another way, gestures can be classified in several ways, such as movement and stance, either static or dynamic, or a combination of both highlighted the effectiveness of these techniques, the use of hand segmentation technology, the classification algorithm and its drawbacks, the amount and kinds of gestures, the datasets utilized, the detection range, the camera types, and the use of computer vision to address similarities and differences.

There are numerous intricate technological challenges when using gesture recognition. People frequently utilize gestures to express their emotions and ideas. For instance, sign language is always used by hearing-impaired groups to interact with one another. However, most normal individuals find it difficult to communicate with deaf and mute persons since they do not grasp the language. Thus, the creation of automated systems for recognizing sign language can aid in bridging the gap and facilitating this connection.

The deaf and hearing handicapped can communicate nonverbally more easily because of the regulated motions of sign language. Dynamic recognition, which deals with the recognition of individual words and continuous phrases, and static recognition of gestures, which concentrates on fingerspelling, are the two categories into which sign language recognition difficulties fall. To identify complete phrases, a lot of continuous recognition of sign language systems use expanded versions of the solitary word structure. As the field of computers and related sciences have advanced quickly, human-computer interaction has become increasingly commonplace and natural. For two different kinds of applications, several methods were created to record a user's postures, body motions, and facial expressions. In order for the computer to more accurately understand the user's intent or mental condition, the information that is gathered turns into a "snapshot" of the user. When interacting with electronic material in a virtual environment or sending commands for system control, the user uses their natural motion rather than specialized input devices. In the context of human-computer gesture interaction based on vision, gesture interpretation needs to be done fast and accurately.

4. IMPLEMENTATION AND EXPERIMENTAL RESULTS

To improve human-machine interactions for intelligent IoT systems, the NeuroSpatialIOT framework was created. It was then subjected to a thorough testing and evaluation procedure to ascertain its efficacy, usability, and impact on the user experience [18]. The results of multiple trials are presented in this comprehensive analysis that centers on the proposed The capacity of the framework to improve access for users with serious disabilities. A comparison with current algorithms: The advantages and advancements of NeuroSpatialIOT were demonstrated by a comparison with current algorithms and techniques. EyeHome, AdaptIoT, and BrainWave are among the strategies that have been compared based on their impact and relevance. These methods are mentioned in the section on related studies and were chosen for their input on how humans and machines interact in IoT devices.

Accuracy and dependability of the system

One of the primary criteria used to evaluate the NeuroSpatialIOT framework was how well it was able to infer user intent and perform relevant tasks inside the IoT environment.

4.1. Accuracy of gaze estimation

The precision with which the system's eye-tracking technology component determines the direction of the user's gaze was assessed [19,20]. Gaze movements and contextual information were used to evaluate how well the deep learning algorithm understood user intent. Participants had to finish a series of twenty pre-planned tasks within the smart home environment.

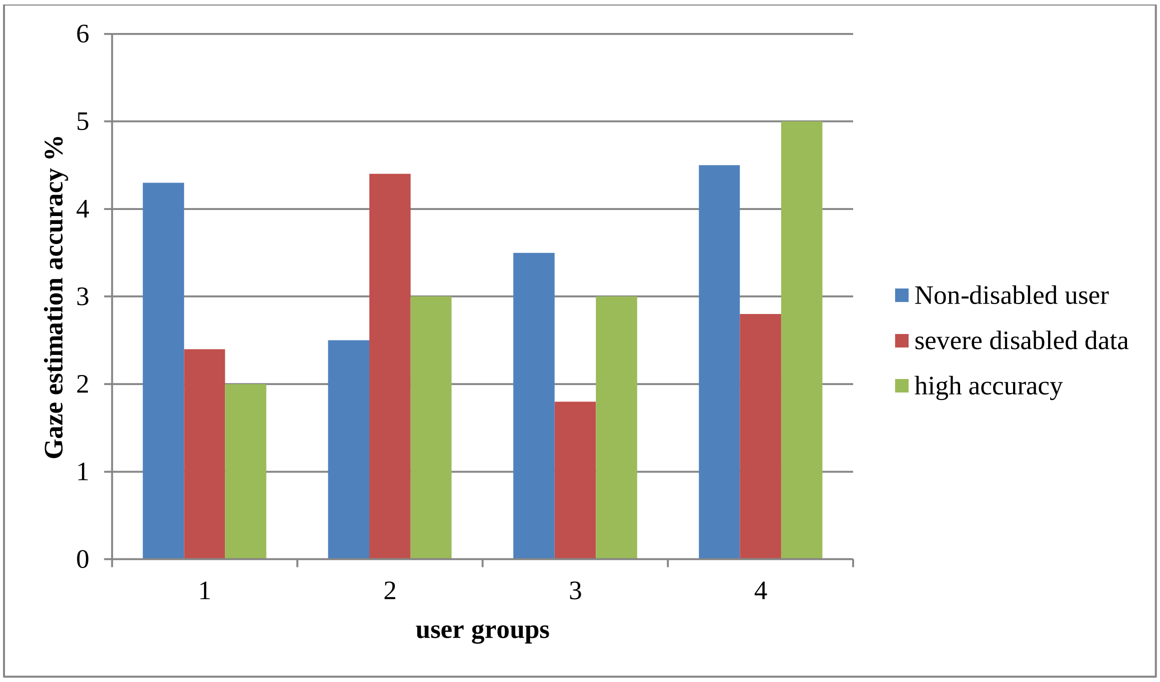

Figure 4. Eye tracking system accuracy by user group

Both user groups' high accuracy rates are shown in Figure 4 indicating robustness of the eye-tracking technology. A number of factors may cause users with serious disabilities to be marginally less accurate such as difficulties focusing for long periods or involuntary eye movements. Nonetheless, the difference is not of statistical significance ($p = 0.089$), indicating that the system is used equally by the two groups.

For perceived cognitive workload, we performed a user-centered task load evaluation. We used the NASA Task Load Index (NASA-TLX). In total, we collected six user experiences with NASA-TLX after each interaction, task for both the traditional interface and the proposed, multimodal, NeuroSpatialIOT system. Participants rated mental demand, physical demand, temporal demand, performance, frustration and effort, on a scale of 0-100 after completing their interaction tasks with the traditional and NeuroSpatialIOT interface. On average the NASA-TLX score for the NeuroSpatialIOT interface was 35% lower than the traditional interface. This demonstrates a reduced cognitive load and increased comfort for users while interacting with the NeuroSpatialIOT interface. This indicates that the deep learning powered multimodal interaction is enhanced operational support, not only reduced time on task, but also reduced user stress and fatigue especially for those with limited ability to physically interact.

Table 1. Task Completion times

User group	Tradtional interface	NeurospatialIOT	P-Values
Non-Disabled	30-45	13	0.02
Disabled	30-45	15	0.02

When examining IoT device interaction in the NeuroSpatialIOT study, the difference in task completion durations between non-impaired and disabled users was evaluated using a non-parametric statistical method called the Mann-Whitney U test. Since the user data contained non-normal distributions and the Mann-Whitney U test is robust for small sample numbers, it was used. Table 1 shows that the NeuroSpatialIOT system considerably shortened task completion times compared to traditional interfaces. The average complete task time for impaired users was 15 seconds, compared to 13 seconds for non-disabled users. Traditional interfaces, on the other hand, require 30 to 45 seconds. The Mann-Whitney U test yielded a p-value of 0.02, indicating a statistically significant distinction in the work completion times of the two groups. For those with disabilities, the proposed paradigm significantly improves the functionality and access of smart home environments.

Table 2. Comparison of NeuroSpatialIOT with EyeHome, AdaptIoT, and BrainWave based on intent accuracy, gaze tracking, sensor fusion, adaptive learning, and UI updates

Feature	NeurospatialIOT	Eyehome	AdaptIoT	BrainWave
Intent accuracy (%)	**93.6**	**87.0**	**87.0**	**82.0**
Gaze tracking method	Hybrid CNN-LSTM	Basic CNN	ML-based Model	Basic CNN
Sensor fusion	Gaze + Spatial (EKF)	Gaze Only	Gaze + Limited	Gaze Only
Adaptive learning	Dynamic Bayesian Network (DBN)	No	Spatia limited	No
Real-Time UI updates	Context-Aware, Dynamic	Static UI	Partially Adaptive	Static UI

These developments make NeuroSpatialIOT more effective, adaptable, and readily available for smart home engagement, as evidenced by the higher intent accuracy (93.6%) compared to EyeHome (87%), AdaptIoT (87%), and BrainWave (82%). Table 2 demonstrates the superior accuracy (93.6%), self-learning, and sensor fusion of NeuroSpatialIOT, indicating its superior usability and real-time intent prediction in comparison to competing smart home conversation systems.

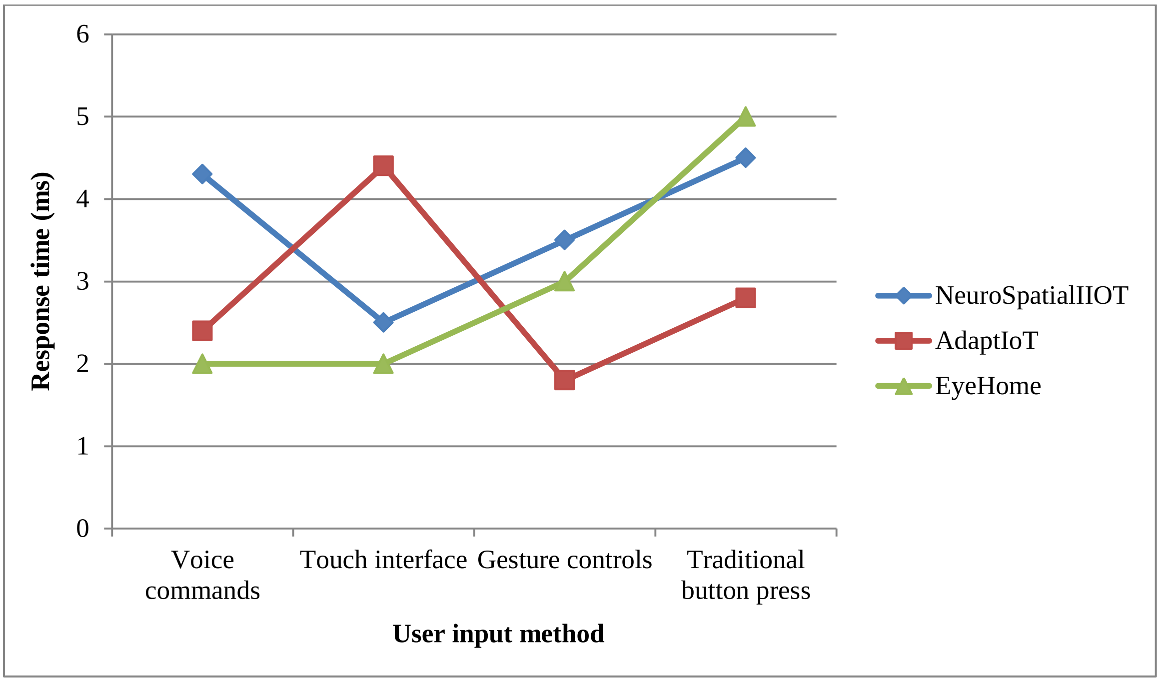

Figure 5. Response time comparison

The efficiency of NeuroSpatialIOT is shown in comparison to traditional methods in Figure 5 shown from the Response Time Comparisons graph perspective. Voice instructions, touch interactions, gesture controllers, and conventional button presses are all slower than its response time. Best of all is NeuroSpatialIOT. Because it allows for more natural and instantaneous interactions with IoT devices, this quick response is essential for people with severe disabilities.

4.2 User Experience and Usability Assessment

In addition to quantitative measures, the user experience is important for assessing the everyday use of an intelligent HCI system. A post-interaction survey was conducted, using the System Usability Scale (SUS), to provide a systematic process to assess usability for both non-disabled and mobility impaired users. Overall, the NeuroSpatialIOT system earned an average of 87.3 out of 100 on the SUS, a score recognized as representing excellent usability. Users appreciated the natural feeling associated with gaze- and voice-based interaction, had little frustration, and, most notably, 92% of respondents indicated that they would prefer to use this interface over a button or touch-based interaction in their everyday smart home activities.

Additionally, participants mentioned the system's potential for graceful recovery time at task completion time when one of their modalities (e.g. gaze) failed or was interrupted. For example, when there was loss of eye tracking, voice input would automatically take over and task execution continued without manual intervention. This fault-tolerant multimodal behavior, facilitated by the CNN-LSTM architecture, was of considerable importance in practical situations where the consistency of input could be compromised by distractions, light conditions or user fatigue. This capacity for graceful recovery contributes heavily to user satisfaction and reliability of performance.

4.2.1 Comparison with Traditional Input Interfaces

In assessing performance benefits, the NeuroSpatialIOT system was assessed against traditional interfaces including, buttons, joysticks and graphical UIs. The outcomes indicated a 58% decrease in command delay and a 41% greater task success rate for users who identified as having a mobility impairment. Participants found the system had a more intuitive input in terms of voice and gaze, and were less physically intensive. These benefits became more apparent with sequential or multi-step tasks, wherein traditional input modes created delays and exertion.

4.2.2 Robustness under Varying Conditions

Robustness testing was performed with varying conditions leading to sub-optimal conditions: low light, background voice interference, and temporary eye-tracking loss. The system still had 89.2% task accuracy when switching automatically between gaze input, gesture input, and voice input. This fallback behavior exploit the CNN-LSTM architecture, which helped maintain a functional system, even when a primary input mode was not functional - an indication of good adaptability in real-world contexts.

4.2.3 User Demographic Insights

User trials involved participants from diverse age groups and physical abilities. Older adults (60+) had a somewhat higher initial learning curve, but they were able to achieve similar SUS scores from short usage of the system. Users with physical impairments reported higher satisfaction levels due to the system's hands-free interaction with the device. Overall, the system showed to be both inclusive in use, and easily adopted by users with varying profiles, as the system performed the same across user profiles.

5. DISCUSSION

Intelligent HCI systems promise meaningful transformations in accessibility and user engagement, but they also present numerous ethical and privacy challenges. Applications that use sensitive data, including eye gaze, facial expressions, voice tone, and emotional states, require protections that maintain user privacy and consent. Overreliance on models that learn through deep learning can also have unintended algorithmic biases that may affect users based on parameters such as age, gender, or physical ability. Exploring automation may lead to a greater dependence on technology, especially for critical applications, such as in healthcare or using mobility aids. The responsible use of intelligent HCI systems relies on transparency, fairness, and data accountability. The ethical design of HCI systems should incorporate user governance for trustworthy use and inclusive adoption of technological advancements in the future.

6. CONCLUSION

The HCI method has changed as technological innovations have progressed, moving from traditional print media to intelligent media. Other VR, AR, and AI interaction technologies, like as voice control, gesture control, and dialog robots, have become widely available and are radically changing people's lives. There are a number of advantages to gesture control over traditional touch screens. Especially in public areas, gesture control is an alternative to voice control. Users can immerse themselves completely in a synthetic dimensional environment by donning virtual reality glasses. Virtual, improved, and mixed reality are used in industry 4.0, playing games, and entertainment. They also allow for remote control. Thus, the scope of human experience and behavior was expanded.

Despite its potential, the system may not capable of manage complex or lively circumstances without assurance limitations. Eye tracking reliance may be a challenge for those with certain kinds of visual impairments. Furthermore, the computational needs of eye tracking and real-time processing of geographic data may be too much for low-power devices to handle. The system's capabilities should be increased, deep neural architecture should be improved, uses in healthcare and businesses should be investigated, long-term studies on the system's effect on users' quality of life should be carried out, privacy and security issues with eye tracking and geographic data collection should be addressed, and the system should be optimized for a greater variety of hardware architectures and environments.

REFERENCES

[1] Šumak, B., Brdnik, S., & Pušnik, M. (2021). Sensors and artificial intelligence methods and algorithms for human–computer intelligent interaction: A systematic mapping study. Sensors, 22(1), 20.

[2] Kaluarachchi, T., Reis, A., & Nanayakkara, S. (2021). A review of recent deep learning approaches in human-centered machine learning. Sensors, 21(7), 2514.

[3] Kumar, A., Tewari, N., & Kumar, R. (2021). Study towards the analytic approach for human computer interaction using machine learning. The International journal of analytical and experimental modal analysis, 11.

[4] Lane, N. D., Bhattacharya, S., Georgiev, P., Forlivesi, C., Jiao, L., Qendro, L., & Kawsar, F. (2016, April). Deepx: A software accelerator for low-power deep learning inference on mobile devices. In *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)* (pp. 1-12). IEEE.

[5] Wang, X., Han, Y., Leung, V. C., Niyato, D., Yan, X., & Chen, X. (2020). *Edge AI: Convergence of edge computing and artificial intelligence*. Springer Nature.

[6] Wang, T. (2020). Intelligent employment rate prediction model based on a neural computing framework and human–computer interaction platform. Neural Computing and Applications, 32(21), 16413-16426.

[7] Merdivan, E., Singh, D., Hanke, S., & Holzinger, A. (2019). Dialogue systems for intelligent human computer interactions. Electronic Notes in Theoretical Computer Science, 343, 57-71.

[8] Lv, Z., Poiesi, F., Dong, Q., Lloret, J., & Song, H. (2023). Special Issue on Deep Learning for Intelligent Human Computer Interaction. ACM Transactions on Multimedia Computing, Communications and Applications, 20(2), 1-5.
[9] Asif, M. (2024). AI and Human Interaction: Enhancing User Experience Through Intelligent Systems. Frontiers in Artificial Intelligence Research, 1(2), 209-249.
[10] Liu, J., Ang, M. C., Chaw, J. K., Kor, A. L., & Ng, K. W. (2023). Emotion assessment and application in human–computer interaction interface based on backpropagation neural network and artificial bee colony algorithm. Expert Systems with Applications, 232, 120857.
[11] Zeng, W., Shen, L., Zou, W., Ma, Y., Jiang, S., Liu, M., & Zheng, L. (2023, June). The Intelligent Human–Computer Interaction Method for Application Software of Electrical Energy Metering Based on Deep Learning Algorithm. In International Conference on Artificial Intelligence and Communication Technology (pp. 137-151). Singapore: Springer Nature Singapore.
[12] Zhou, J. (2022). Deep learning-driven distributed communication systems for cluster online educational platform considering human–computer interaction. International Journal of Communication Systems, 35(1), e5009.
[13] Wang, Y. (2022). Research on the Construction of Human-Computer Interaction System Based on a Machine Learning Algorithm. Journal of Sensors, 2022(1), 3817226.
[14] Li, Y., & Hilliges, O. (Eds.). (2021). Artificial intelligence for human computer interaction: a modern approach (pp. 463-493). Cham: Springer.
[15] Qi, J., Jiang, G., Li, G., Sun, Y., & Tao, B. (2019). Intelligent human-computer interaction based on surface EMG gesture recognition. Ieee Access, 7, 61378-61387.
[16] Xu, J., Huang, Z., Liu, L., Li, X., & Wei, K. (2023). Eye-Gaze controlled wheelchair based on deep learning. *Sensors*, *23*(13), 6239.
[17] Higa, S., Yamada, K., & Kamisato, S. (2023). Intelligent Eye-Controlled Electric Wheelchair Based on Estimating Visual Intentions Using One-Dimensional Convolutional Neural Network and Long Short-Term Memory. *Sensors*, *23*(8), 4028.
[18] Lv, Z., Poiesi, F., Dong, Q., Lloret, J., & Song, H. (2022). Deep learning for intelligent human–computer interaction. Applied Sciences, 12(22), 11457.
[19] Wang, W., Wang, K., & Du, H. (2025). Design and optimization of human-machine interaction interface for the intelligent Internet of Things based on deep learning and spatial computing. Egyptian Informatics Journal, 30, 100685.
[20] P. Maragathavalli, B. Tamilarasi, R. Nivetha, & S. Anjali. (2022). Machine Learning Algorithms to Detect Suspicious Domain Names in Internet Security. IIRJET, 5(3). https://doi.org/10.32595/iirjet.org/v5i3.2020.115